REMARKS/ARGUMENTS

The Examiner is thanked for the review of the application.

Claims 1-14 remain in this application. Claims 1, 3, 4, 8, 9, 12 and 13 have been amended. Claims 15-20 have been added. Support for added Claims 15, 16 and 17 can be found in page 99, line 8 to page 103, line 4 of the Specification as filed. Support for added Claims 18, 19 and 20 can be found in page 60, line 5 to page 73, line 8 of the Specification as filed.

In the Office Action dated August 31, 2006, the Examiner has objected the Specification stating that "the 'Attorney Docket Number DEM1P002)' should be deleted and application no. 09/741,956 should be entered for the application filed December 20, 2000 entitled 'Econometric Engine', by Hau Lee, Suzanne Valentine, ...';09/741,958 filed December 20, 2000 entitled 'Price Optimization System', by Michael Neal, ...' and 09/741,959 filed December 20, 2000, entitled 'Econometric Optomization Engine', by Krishna Venkatraman, ...'." In the Specification, the paragraph starting at page 1 beginning at line 17, paragraph starting at page 2 beginning at line 3, and paragraph staring at page 2 beginning at line 8 have been amended to comply with MPEP 608.01(b).

The specification has been amended accordingly. All references to "Attorney Docket Numbers" have been removed. Also, the status (Pending or Allowed) of each cross-referenced application has been added.

Also, in the Office Action dated August 31, 2006, the Examiner has rejected Claims 1, 3, 5 and 6 as having terminology used inconsistently with accepted meaning, stating that "[w]here Applicants' act as their own lexicographer to specifically define a term of a claim contrary to its ordinary meaning, the written description must clearly redefine the claim term and set forth the uncommon definition so as to put one reasonably skilled in the art on notice that the Applicants' intended to so redefine that claim term. Process Control Corp. v. HydReclaim Corp., 190 F.3d

1350, 1357, 52 USPQ2d 1029, 1033 (Fed. Cir. 1999). The term '[imputed]' in claims 1, 3, 5, and 6 is confusing because the accepted meaning is 'to lay responsibility or blame for often falsely or unjustly, and to credit a person or a cause'. Merriam-Webster's Collegiate Dictionary - 10 ED. The term is indefinite because the specification does not clearly redefine the term."

Applicants respectfully refer to the Information Discolsure Statement filed October 12, 2005, which includes a chapter discussing imputation from a textbook entitled "Bayesian Data Analysis" by Gelman, Carlin, Stern and Rubin, and also includes Webster's definitions for "infer". Additional copies of the material have been enclosed with this amendment for the Examiner's convenience.

On page 453 of the Gelman et al, sub-chapter 17.7 titled "Inference using multiple imputation", the authors describe how imputation can be used in complex surveys to provide a complete dataset by replacing each of the missing values by an imputed value when some data values are missing. Applicants also submit the Webster's definitions of "infer", inferred", "inferring" and "inference" which is also consistent with Applicant's use of "imputed" in this application.

Accordingly, the paragraph at page 17 beginning at line 16 has been amended to incorporate this commonly used mathematical definition and hence definitively points out the claimed invention. Applicants would further like to draw the Examiner's attention to lines of page of the specification as filed, which states "A further advantageous aspect of the present invention is that, even if such size or UOM information is incomplete or not provided, it can also be imputed", which most mathematicians skilled in the art will understand as synonymous to "can be inferred".

Also, in the Office Action dated August 31, 2006, the Examiner has rejected Claims 1 and 3 as having terminology used inconsistently with accepted meaning, stating that "[t]he term '[posterior]' inference in claims 1 and 3 is confusing because the accepted meaning is 'coming after, situated behind'. Merriam-Webster's Collegiate Dictionary -10 ED. The term is indefinite because the specification does not clearly redefine the term."

Base Claim 1 has been amended to include "generating imputed variables, suing the computer system, wherein said imputed variables are generated by imputing at least one missing data point when the at least one data point is missing." Support may be found in page 23, line 9

to page 28, line 5 of the Specification as filed, which states "Next the process includes determining the nature of missing data records in a fourth error detection and correction step ..."

Base Claim 3 has been amended to include "an econometric engine for receiving sales data from at least one store via the network, cleansing the sales data and generating imputed variables, wherein said imputed variables are generated by imputing at least one missing data point when the at least one data point is missing". Support may be found in page 23, line 9 to page 28, line 5 of the Specification as filed, which states "Next the process includes determining the nature of missing data records in a fourth error detection and correction step ..."

Additionally, in the Office Action dated August 31, 2006, the Examiner has rejected Claims 1-14 under 35 U.S.C. 112, second paragraph, stating that claims are " indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Claim 1 recites 'receiving sales data' and receiving cost data'. It is vague and unclear how 'sales data' and the 'cost data' are received and what is used to generate the 'imputed variables'. Are the 'imputed variables' generated by machine or a user and is the 'sales data' and the 'cost data' received from a machine or a user? Claim 3 recites 'ecometric engine'. It is unclear and vague what the 'ecometric engine' is from the Specification and the drawings. 'Barron's Dictionary of Computer and Internet Terms' on page 127 defines the term 'engine' as 'the part of a computer program that implements a special technique'. The dependent claims 2 and 4-14 are also rejected because of their dependency on a rejected base claim."

Claim 1 has now been amended to recite, in relevant part:

"In a computer system, a computer-implemented method for modeling cost, useful in association with at least one store and an optimization engine coupled to the computer system via a network, the computer-implemented method comprising: receiving sales data, using the computer system, from the at least one store via the network; cleaning the sales data, using the computer system; generating imputed variables, using the computer system, wherein said imputed variables are generated by imputing at least one missing data point when the at least one data point is missing; receiving cost data, using the computer system, from the at least one store via the network; estimating cost per unit of product, using the computer system, from the sales

data, the imputed variables and the cost data; and outputting the estimated cost per unit of product, using the computer system, to the optimization engine via the network." (emphasis added).

Support for the amendment may be found on page 10, lines 5-15 of the Specification as filed, which states "The financial model engine 108 may receive data 132 from the stores 124 ... Cost data 136 is provided from the financial model engine 108 to the optimization engine 112 (step 224)." Additional Support may be found on page 14, lines 10-17 of the Specification as filed, which states "The raw econometric data must specify the store from which the data is collected... and product size." Additional Support may be found on page 75, lines 9-17 of the Specification as filed, which states "In a preferred embodiment of the invention, the stores may only need to supply ... using these estimations, costs may be more easily calculated on a store level, instead of being averaged over all of the stores." Additional Support may be found on page 116, lines 4-12 of the Specification as filed, which states "FIG'S. 7A and 7B illustrate a computer system 900, which forms part of the network 10 and is suitable for implementing embodiments of the present invention..."

Claim 3 has now been amended to recite, in relevant part:

"An apparatus for modeling costs, useful in association with at least one store coupled to the apparatus via a network, and useful in association with an optimization engine, wherein the optimization engine is configured to receive input from the apparatus, and wherein the optimization engine is further configured to generate a preferred set of prices, the apparatus comprising: an econometric engine for receiving sales data from at least one store via the network, cleansing the sales data and generating imputed variables, wherein said imputed variables are generated by imputing at least one missing data point when the at least one data point is missing; and a financial engine for receiving imputed variables from the econometric engine, receiving cost data from at least one store via the network, generating a cost model, and outputting the cost model to the optimization engine." (emphasis added).

Support for the amendment may be found on page 10, lines 5-9 of the Specification as filed, which states "The financial model engine 108 may receive data 132 from the stores 124 ... and fixed cost data." Additional Support may be found on page 14, lines 10-17 of the

Specification as filed, which states "The raw econometric data must specify the store from which the data is collected... and product size." Additional Support may be found on page 75, lines 9-17 of the Specification as filed, which states "In a preferred embodiment of the invention, the stores may only need to supply ... using these estimations, costs may be more easily calculated on a store level, instead of being averaged over all of the stores." Additional Support may be found on Figures 1 and 5 of the Specification as filed.

Claim 4 has now been amended to include "the financial engine estimates inventory space in the at least one store used by a product from the sales data and delivery data." Support for the amendment may be found on page 75, lines 13-15 of the Specification as filed, which states "may infer the amount of shelf-space an item utilizes from the cubic feet of the item, the volume of sales of the item, and how often the store is replenished with the item."

Claim 8 has now been amended to include "said estimating cost are estimated for each of the at least one store." Support for the amendment may be found on page 75, lines 18-19 of the Specification as filed, which states "The tailoring of costs per store allows the maximization of profits for each store."

Claim 9 has now been amended to include "said estimated cost per unit of product is determined as a cost for said product in said each of the at least one store for a selected demand group in a selected time period, further wherein said demand group is a group of highly substitutable products." Support for the amendment may be found on page 75, lines 3-5 of the Specification as filed, which states "The financial model engine ... cost for a particular product (k) given a store (s), demand group (i), and a day (t)."

Claim 12 has now been amended to include "said cost model models costs for each of the at least one store." Support for the amendment may be found on page 75, lines 18-19 of the Specification as filed, which states "The tailoring of costs per store allows the maximization of profits for each store."

Claim 13 has now been amended to include "said cost model models costs for individual products in said each of the at least one store for a selected demand group in a selected time period, further wherein said demand group is a group of highly substitutable products." Support for the amendment may be found on page 75, lines 3-5 of the Specification as filed, which states "The financial model engine ... cost for a particular product (*k*) given a store (*s*), demand group (*i*), and a day (t)."

New Dependent claim 15 includes "the econometric engine is coupled to a coefficient estimator, wherein the coefficient estimator generates a combined product sales model, a share model and a sales model." Support for the amendment may be found on page 12, lines 3-5 of the Specification as filed, which states "The econometric engine comprises an imputed variable generator 304 and a coefficient estimator 308." Additional Support for the amendment may be found on page 60, lines 6-13 of the Specification as filed, which states "The coefficient estimator 308 uses... sales for a demand group (S) is calculated and a market share (F) for a particular product is calculated, so that demand (D) for a particular product is estimated by D=S•F..."

New Dependent claim 16 includes "the coefficient estimator outputs the combined product sales model to the optimization engine, and wherein the optimization engine generates optimized pricing for the products from the combined product sales model and cost model." Support can be found in page 98, lines 13-17 of the Specification as filed, which states "The optimization engine uses the group sales equation and the market share equation previously defined in the section on the econometric engine to predict group sales and product market share, respectively. These two are then combined to predict product sales at the store level."

New Dependent claim 17 includes "the coefficient estimator receives imputed variables from the econometric engine and sales data from the at least one store." Support can be found in page 60, lines 6-7 of the Specification as filed, which states "The coefficient estimator 308 uses the imputed variables and data to estimate coefficients, which may be used in an equation to predict demand."

New Dependent claim 18 includes "the combined product sales model is given by:

$$\hat{D}_{i,k,t} = \hat{F}_{i,k,t}\hat{S}_{i,t}$$

wherein,

k = a product

i = a primary demand group

t = a time period

$D_{i,k,t}$ = a demand for product k in demand group i in time period t

$F_{i,k,t}$ = a fraction of the demand group i equivalent sales comprised by the product k in the time period t

$S_{i,t}$ = an equivalent sales of the demand group i in the period t.

Support can be found in page 60, line 5 to page 73, line 8 of the Specification as filed, which states "The coefficient estimator 308 uses the imputed variables and data to estimate coefficients, which may be used in an equation to predict demand ..."

New Dependent claim 19 includes "the sales model is given by:

$$\left(\frac{\hat{S}_{i,t}}{S_{Bi,t}}\right) = \exp\left(\begin{array}{l}\hat{\mathrm{K}}_i + \hat{\gamma}_i \frac{P_{i,t}}{\overline{P}_{i,t}} + \hat{\nu}_i M_{i,t} + \hat{\psi}_i X_{i,t} + \hat{\kappa}_i X_{i,t}\frac{P_{i,t}}{\overline{P}_{i,t}} + \sum_{n=1}^{\tau}\hat{\delta}_{i,n}\frac{\sum_{r=t-mn}^{t-m(n-1)-1} S_{i,r}}{\sum_{r=t-mn}^{t-m(n-1)-1}\overline{S}_{i,r}} + \sum_{j\neq i}\hat{\phi}_{i,j}\frac{\hat{S}_{j,t}}{\overline{S}_{j,t}} \\ + \hat{\eta}_{i,t}\left(\frac{\overline{P}_{i,t} - \overline{\overline{P}}_{i,t}}{\overline{\overline{P}}_{i,t}}\right) + \hat{\pi}_i\frac{TS_t}{T\overline{S}_t} + \hat{\theta}_i\frac{S_{i,t-7}}{\overline{S}_{i,t-7}} + \frac{\hat{\sigma}^2}{2}\end{array}\right)$$

wherein,

k = the product

i = the primary demand group

j = a secondary demand group

t = the time period

B = a baseline state of product

$S_{i,t}$ = the equivalent sales of the demand group i in the period t

$S_{Bi,t}$ = an equivalent baseline sales of the demand group i in the period t

TS_t = total sales for the store in the period t

$T\overline{S}_t$ = total sales for a region in the period t

$P_{i,t}$ = an equivalent price of the demand group i in the time period t

$\overline{P}_{i,t}$ = an average equivalent price of the demand group i for the time period t

$\overline{\overline{P}}_{i,t}$ = an average competitor equivalent price of the demand group i for the time period t

$M_{i,t}$ = a promotion level for the demand group i in the time period t

$X_{i,t}$ = a seasonality index for the demand group i in the time period t

γ_i = a price elasticity factor for the demand group i

ν_i = a promotion factor for the demand group i

ψ_i = a seasonality factor for the demand group i

κ_i = a seasonality-price interaction factor that measures the interaction of weighted average price deviations and seasonality for the demand group i

n = a number of time periods away from the time period t

$\delta_{i,n}$ = a time lag factor for the demand group i and the delay of n weeks

$\phi_{i,j}$ = a cross elasticity factor for the demand group i and the demand group j

$\eta_{i,t}$ = a competitive price factor for the demand group i measured with respect to the difference between the weighted average price of the demand group within the store and outside competitors

π_i = a traffic factor for the demand group i

θ_i = a day-of-week effect for the demand group i

$\hat{\sigma}^2$ = a mean square error of the sales model divided by 2

K_i = a constant associated with the demand group i

Support can be found in page 60, line 5 to page 73, line 8 of the Specification as filed, which states "The coefficient estimator 308 uses the imputed variables and data to estimate coefficients, which may be used in an equation to predict demand ..."

New Dependent claim 20 includes "the share model is given by:

$$\hat{F}_{i,k,t} = \frac{\exp\left\{\hat{\Lambda}_{i,k} + \hat{\rho}_{i,k}\left(P_{Ri,k,t}\right) + \sum_{p=1}^{n_p}\hat{\sigma}_{p,i,k}\left(M_{p,i,k,t}\right) + \sum_{n=1}^{\tau}\hat{\chi}_{i,k,n}\sum_{r=t-mn}^{t-m(n-1)-1}\left(F_{i,k,r}\right)\right\}}{\sum_{l\in Dem_i}\exp\left\{\hat{\Lambda}_{i,l} + \hat{\rho}_{i,l}\left(P_{Ri,l,t}\right) + \sum_{p=1}^{n_p}\hat{\sigma}_{p,i,l}\left(M_{p,i,l,t}\right) + \sum_{n=1}^{\tau}\hat{\chi}_{i,k,n}\sum_{r=t-mn}^{t-m(n-1)-1}\left(F_{i,l,r}\right)\right\}}$$

wherein,

k = the product

i = the primary demand group

t = the time period

n = the number of time periods away from the time period t

$F_{i,k,t}$ = the fraction of the demand group i equivalent sales comprised by the product k in the time period t

$P_{Bi,k,t}$ = an equivalent base price of the product k in the demand group i in the time period t

$\overline{P}_{Bi,(k),t}$ = an average equivalent base price of all products other than the product k in the demand group i for the time period t

$P_{RBi,k,t}$ = a relative equivalent base price of the product k in the demand group i for the time period t

$\overline{P}_{RBi,\bullet,t}$ = an average relative equivalent base price in the demand group i for the time period t

$M_{p,i,k,t}$ = a level of promotion type p for the product k in the demand group i in the time period t

$\rho_{i,k}$ = a relative base price elasticity factor for the product k in the demand group i

$\sigma_{p,i,k}$ = a promotion factor p for the product k in the demand group i

$\chi_{i,k,n}$ = a time lag factor for the product k in the demand group i and the delay of n

$\Lambda_{i,k}$ = a constant associated with the product k in the demand group i

Support can be found in page 60, line 5 to page 73, line 8 of the Specification as filed, which states "The coefficient estimator 308 uses the imputed variables and data to estimate coefficients, which may be used in an equation to predict demand ..."

It is respectfully submitted that amended base claim 1 and 3 particularly point out and distinctly claim the subject matter regarded as the invention. New Claims 15-20 have been added refining the subject matter of the invention. Furthermore, Claims 2, 4-20 all depend on independent Claims 1 and 3, and are also allowable over the cited art for at least the same reasons.

In sum, base Claims 1 and 3 have been amended and are believed to be allowable. Dependent Claims 4, 8, 9, 12 and 13 have been amended and are believed to be allowable. Dependent Claims 2, 4-20 which depend therefrom are also believed to be allowable as being dependent from their patentable parent Claims 1 and 3 for at least the same reasons. Hence, Examiner's rejections of dependent claims 2, 4-14 are rendered moot in view of independent Claims 1 and 3. Applicants believe that all pending Claims 1-20 are now allowable over the cited art and are also in allowable form and respectfully request a Notice of Allowance for this application from the Examiner. The commissioner is authorized to charge any fees that may be due to our Deposit Account No. 50-2766 (Order No. DEM1P004). Should the Examiner believe that a telephone conference would expedite the prosecution of this application, the undersigned can be reached at telephone number 925-570-8198.

LAW OFFICES OF KANG S. LIM
PMB 436
3494 Camino Tassajara Road
Danville, CA 94506
Voice: (925) 570 8198
Facsimile: (925) 736 3974

Respectfully submitted,

/Kang S. Lim/

Kang S. Lim
Attorney for Applicant(s)
Reg. No. 37,491

CUSTOMER NO. 36088

\\KSL IDS d

SECOND COLLEGE EDITION

WEBSTER'S NEW WORLD DICTIONARY

of the American Language

David B. Guralnik, *Editor in Chief*

THE WORLD PUBLISHING COMPANY

New York and Cleveland

BEST AVAILABLE COPY

WEBSTER'S NEW WORLD DICTIONARY, Second College Edition

Library of Congress Catalog Card Number: 71-182408

PRINTED IN THE UNITED STATES OF AMERICA

disgrace; dishonor 2. the quality of being ickedness 3. an infamous act 4. *Law* loss of certain civil rights sustained by a of an infamous crime —*SYN.* see DIS-

(ē) *n., pl.* -cies [LME. < L. *infantia*] 1. of being an infant; babyhood; very early eginning or earliest stage of anything 3. eriod of being a minor; period before the ity, usually twenty-one; minority [ME. *infaunt* < OFr. *enfant* < L. *infans* ld < *adj.*, not yet speaking < *in-*, not + to speak: see FAME] 1. a very young erson in the state of legal infancy; minor nfants or infancy 2. in a very early stage , -fän'-) *n.* [Sp. & Port., fem. of *infante*: ughter of a king of Spain or Portugal nfante

ip. & Port. < L. *infans*: see INFANT] any pain or Portugal, except the heir to the

an'tə sīd') *n.* [Fr. < LL. *infanticidium* < ıo kills an infant: see INFANT & -CIDE] oaby 2. [Fr. < LL. *infanticida*] a person

til', -tīl) *adj.* [L. *infantilis*] 1. of or nfants or infancy 2. like, suitable for, or ı infant; babyish; childish or childlike; ; earliest stage of development

same as POLIOMYELITIS

an t'l iz'm; *esp. in psychol.*, in fan'-) *n.* ih behavior; specif., *Psychol.* an abnor- such behavior persists into adult life; arded mental and physical growth and e sexually

tin', -tin) *adj.* [Fr. *infantin, enfantin*] childish or childlike

rē) *n., pl.* -tries [Fr. *infanterie* < It. very young person, knight's page, foot : see INFANT] 1. foot soldiers collec- anch of an army consisting of soldiers ed to fight chiefly on foot 2. [I-] a y regiment [the 274th *Infantry*]

ıən) *n., pl.* -men (-mən) a soldier in the

. [ML. *infarctus*, for L. *infartus*, pp. of *farcire*, to stuff: see FARCE] an area of resulting from obstruction of the blood plying the part

:'shən) *n.* 1. the development of an NFARCT

) *n.* [ME. *infer*, entrance < OE. *infær* going < *faran*: see FARE] [Dial.] a party after a wedding, usually on the

ı'ōo wāt') *vt.* -at'ed, -at'ing [< L. *atuare*, to make a fool of < *in-*, intens. FATUOUS] 1. to make foolish; cause to : 2. to inspire with foolish or shallow *adj.*-infatuated —*n.* a person who is

adj. 1. lacking sound judgment; y carried away by foolish or shallow ı·fat'u·at'ed·ly *adv.*

ıch'ōo wā'shən) *n.* [LL. *infatuatio*] an infatuated —*SYN.* see LOVE

ı b'l) *adj.* not feasible; not easily done; ıa'si·bil'i·ty *n.*

[ME. *infecten* < MFr. *infecter* < L. *e*, to put or dip into, tinge, stain < *in-*, 1. to contaminate with a disease- or matter 2. to cause to become dis- to contact with such an organism or an individual, organ, tissue, etc.: said iism 4. to affect or imbue with one's ip. in a harmful or undesirable way

hən) *n.* [ME. *infeccioun* < OFr. *io*] 1. an infecting; specif., *a*) the act diseased *b*) the act of affecting with s 2. the fact or state of being infected, in the body of bacteria, protozoans, sites 3. something that results from ected; specif., *a*) a disease resulting 2) *b*) a feeling, belief, influence, etc. person or another 4. anything that

tj. 1. likely to cause infection; con- cing organisms or matter 2. desig- can be communicated by infection o spread or to affect others; catching 4. [Obs.] infected with disease —*in-* *fec'tious·ness n.*

a viral disease causing inflammation cterized by jaundice, fever, etc.

eosis an acute disease, esp. of young by fever, swollen lymph nodes, sore ities of the lymphocytes: it is pre- y a virus

in·fec·tive (in fek'tiv) *adj.* [ME. *infectif* < OFr. < L. *infectivus*] likely to cause infection; infectious

in·fe·cund (in fē'kənd, -fek'ənd) *adj.* [ME. *infecunde* < L. *infecundus*] not fecund; not fertile; barren —**in·fe·cun·di·ty** (in'fi kun'də tē) *n.*

in·fe·lic·i·tous (in'fə lis'ə təs) *adj.* not felicitous; unfortunate or unsuitable —**in'fe·lic'i·tous·ly** *adv.*

in·fe·lic·i·ty (-tē) *n.* [L. *infelicitas* < *infelix*, unfortunate: see IN-[3] & FELICITY] 1. the quality or condition of being infelicitous 2. *pl.* -ties something infelicitous; unsuitable or inapt remark, action, etc.

in·fer (in fur') *vt.* -ferred', -fer'ring [L. *inferre*, to bring or carry in, infer < *in-*, in + *ferre*, to carry, BEAR[1]] 1. orig., to bring on or about; cause; induce 2. to conclude or decide from something known or assumed; derive by reasoning; draw as a conclusion 3. *a*) to lead to as a conclusion; indicate *b*) to indicate indirectly; imply: in this sense, still sometimes regarded as a loose usage —*vi.* to draw inferences —**in·fer'a·ble** *adj.* —**in·fer'a·bly** *adv.* —**in·fer'rer** *n.*

SYN.—**infer** suggests the arriving at a decision or opinion by reasoning from known facts or evidence [from your smile, I *infer* that you're pleased]; **deduce**, in strict discrimination, implies inference from a general principle by logical reasoning [the method was *deduced* from earlier experiments]; **conclude** strictly implies an inference that is the final logical result in a process of reasoning [I must, therefore, *conclude* that you are wrong]; **judge** stresses the careful checking and weighing of premises, etc. in arriving at a conclusion; **gather** is an informal substitute for **infer** or **conclude** [I *gather* that you don't care]

in·fer·ence (in'fər əns) *n.* [ML. *inferentia*] 1. the act or process of inferring; specif., the deriving of a conclusion in logic by either induction or deduction 2. something inferred; specif., a conclusion arrived at in logic

in·fer·en·tial (in'fə ren'shəl) *adj.* [< ML. *inferentia* + -AL] based on or having to do with inference —**in'fer·en'tial·ly** *adv.*

in·fe·ri·or (in fir'ē ər) *adj.* [ME. < L., compar. of *inferus*, low, below < IE. **ndheros*, whence UNDER] 1. lower in space; placed lower down 2. low or lower in order, status, rank, etc.; subordinate 3. lower in quality or value than (with *to*) 4. poor in quality; below average 5. *Anat.* located below or directed downward 6. *Astron.* between the earth and the sun [Mercury and Venus are *inferior* planets] 7. *Bot.* having the sepals, petals, and stamens attached at the apex: said of the ovary of an epigynous flower 8. *Printing* placed below the type line, as 2 in NO_2 —*n.* an inferior person or thing —**in·fe'ri·or'i·ty** (-ôr'ə tē, -är'-) *n.*

inferiority complex 1. *Psychol.* a neurotic condition resulting from various feelings of inferiority, such as derive from real or imagined physical or social inadequacy and often manifested through overcompensation in excessive aggressiveness, a domineering attitude, etc. 2. popularly, any feeling of inferiority, inadequacy, etc.: cf. SUPERIORITY COMPLEX

in·fer·nal (in fur'n'l) *adj.* [ME. < OFr. < LL. *infernalis* < L. *infernus*, underground, lower, infernal < *inferus*: see INFERIOR] 1. *a*) of the ancient mythological world of the dead *b*) of hell 2. hellish; diabolical; fiendish; inhuman 3. [Colloq.] hateful; outrageous —**in·fer'nal·ly** *adv.*

infernal machine *earlier name for* a booby trap or time bomb

In·fer·no (in fur'nō) *n., pl.* -nos [It. < L. *infernus*: see INFERNAL] hell or any place suggesting hell, usually characterized by great heat or flames —[I-] that section of Dante's *Divine Comedy* which describes hell and the sufferings of the damned

in·fe·ro- (in'fə rō') [< L. *inferus*: see INFERIOR] *a combining form meaning* below and *[inferoanterior]*

in·fe·ro·an·te·ri·or (in'fə rō'an tir'ē ər) *adj.* [prec. + ANTERIOR] lying below and in front

in·fer·tile (in fur't'l; *chiefly Brit.*, -tīl) *adj.* [MFr. < L. *infertilis*] 1. not fertile; not productive; barren 2. not fertilized, as an egg —*SYN.* see STERILE —**in·fer·til·i·ty** (in'fur til'ə tē) *n.*

in·fest (in fest') *vt.* [Fr. *infester* < L. *infestare*, to attack, trouble < *infestus*, hostile < *in-*, in + IE. base **dhers-*, to be bold, attack, whence DARE] 1. to overrun or inhabit in large numbers, usually so as to be harmful or bothersome; swarm in or over 2. to be parasitic in or on —**in'fes·ta'tion** *n.* —**in·fest'er** *n.*

in·feu·da·tion (in'fyoo dā'shən) *n.* [M.L. *infeudatio* < pp. of *infeudare*, to enfeoff < *in-*, in + *feudum*: for IE. base see FEE] in feudal law, the granting of an estate in fee; enfeoffment

in·fi·del (in'fə d'l) *n.* [ME. < MFr. *infidèle* < L. *infidelis*, unfaithful (in LL.(Ec.), unbelieving) < *in-*, IN-[1] + *fidelis*: see FIDELITY] 1. a person who does not believe in a particular religion, esp. the prevailing religion; specif., *a*) among Christians, a non-Christian *b*) among Moslems, a non-Moslem 2. a person who holds no religious belief 3. a person who does not accept some particular theory, belief, etc. —*adj.* 1. that is an infidel; unbelieving 2. of infidels —*SYN.* see ATHEIST

in·fi·del·i·ty (in'fə del'ə tē) *n., pl.* -ties [L. *infidelitas* < *infidelis*: see prec.] 1. the fact or state of being an infidel 2. unfaithfulness or disloyalty to another; esp., sexual unfaithfulness of a husband or wife; adultery 3. an unfaithful or disloyal act

in·field (in'fēld') *n.* [IN, *adv.* + FIELD] 1. the land of a farm nearest the farmhouse ☆2. *a*) the area enclosed by the four base lines on a baseball field *b*) the infielders collectively, or the area covered by them Distinguished from OUTFIELD ☆3. the area inside a race track or running track: field events of a track meet are often held on it

☆**in·field·er** (-ər) *n. Baseball* a player whose position is in the infield; shortstop, first baseman, second baseman, or third baseman; the pitcher and the catcher are considered infielders when fielding the ball

in·fight·ing (in'fīt'iŋ) *n.* 1. fighting, esp. boxing at close range 2. intense competition or conflict, often bitterly personal, as between political opponents or within an organization or group —**in'fight'er** *n.*

in·fil·trate (in fil'trāt, in'fil trāt') *vi., vt.* -trat·ed, -trating [IN-[1] + FILTRATE] 1. to pass, or cause (a fluid) to pass, through small gaps or openings; filter 2. to pass through, as in filtering 3. to pass, or cause (individual troops) to pass, through weak places in the enemy's lines in order to attack the enemy's flanks or rear 4. to penetrate, or cause to penetrate, (a region or group) gradually or stealthily, so as to attack or to seize control from within —*n.* something that infiltrates —**in'fil·tra'tion** *n.* —**in'fil·tra'tive** *adj.* —**in'fil·tra'tor** *n.*

infin. infinitive

in·fi·nite (in'fə nit) *adj.* [ME. < L. *infinitus*: see IN-[1] & FINITE] 1. lacking limits or bounds; extending beyond measure or comprehension; without beginning or end; endless 2. very great; vast; immense 3. *a*) *Math.* indefinitely large; greater than any finite number however large *b*) capable of being put into one-to-one correspondence with a part of itself *[infinite set]* —*n.* something infinite, as space or time —**the Infinite (Being)** God —**in'fi·nite·ly** *adv.* —**in'fi·nite·ness** *n.*

in·fin·i·tes·i·mal (in'fin ə tes'ə məl, in fin'-) *adj.* [ModL. *infinitesimus* < L. *infinitus* (see INFINITE), after *centesimus*, hundredth < *centum*, one hundred] 1. too small to be measured; infinitely small 2. *Math.* of or pertaining to an infinitesimal —*n.* 1. an infinitesimal quantity 2. *Math.* a variable nonzero number which is understood to be smaller in absolute value than any preassigned nonzero number —**in'fin·i·tes'i·mal·ly** *adv.*

infinitesimal calculus the methods of mathematical analysis of differential and integral calculus

in·fin·i·tive (in fin'ə tiv) *adj.* [LL. *infinitivus* < L. *infinitivus* (*modus*), lit., unlimited (mood) < *infinitus* (see INFINITE): so named because it is not limited to any person, number, or tense] *Gram.* of or connected with an infinitive —*n. Gram.* the form of the verb which expresses existence or action without reference to person, number, or tense and can also function as a noun: in English, it is usually the form of the first person singular present preceded by the marker *to [to go, to think]* or by another verb form *[can he go, make him try]* —**in·fin'i·ti'val** (-tī'vəl) *adj.*

in·fin·i·tude (-tōōd', -tyōōd') *n.* [< L. *infinitus*, INFINITE, prob. after MAGNITUDE] 1. the quality of being infinite 2. an infinite quantity, number, or extent

in·fin·i·ty (-tē) *n., pl.* -ties [ME. *infinite* < OFr. *infinité* < L. *infinitas*] 1. the quality of being infinite 2. anything infinite; endless or unlimited space, time, distance, quantity, etc. 3. an indefinitely large number or amount 4. *Geom.* a location infinitely distant from the origin or base point of the space being considered 5. *Math.* the numerical value of an infinite quantity: symbol, ∞ 6. *Photog. a*) a distance so far from a camera that rays of light reflected from a subject there may be regarded as parallel *b*) a setting for such a distance on a camera's focusing scale —**to infinity** without limit or end

in·firm (in furm') *adj.* [ME. < L. *infirmus*] 1. not firm or strong physically; weak; feeble, as from old age 2. not firm in mind or purpose; not resolute; vacillating 3. not stable; firm, or sound; frail; shaky, as a structure 4. not secure or valid *[an infirm* title to property*]* —*SYN.* see WEAK —**in·firm'ly** *adv.* —**in·firm'ness** *n.*

in·fir·ma·ry (in fur'mə rē) *n., pl.* -ries [ML. *infirmaria*, *infirmarium* < L. *infirmus*] a place for the care of the sick, injured, or infirm; esp., a building or room, as in a school, that serves as a hospital or dispensary

in·fir·mi·ty (-mə tē) *n.* [ME. *infirmite* < L. *infirmitas*] 1. the quality or state of being infirm; feebleness; weakness 2. *pl.* -ties an instance of this; specif., *a*) a physical weakness or defect; frailty or ailment, as from old age *b*) a moral weakness; defect

in·fix (in fiks'; *also, and for n. always*, in'fiks') *vt.* [< L. *infixus*, pp. of *infigere*, to fix or drive in: see IN-[1] & FIX] 1. to fasten or set firmly in or on, esp. by inserting or piercing 2. to fix firmly in the mind; instill; implant 3. to place (an infix) within the body of a word —*n. Linguis.* a bound morpheme that is added within a word that has no affixes or within that part of a word to which affixes are added (Ex.: Arabic *-ta-* in *iq-ta-riba*, to cause oneself to come near < *qariba*, to come near)

CHAPMAN & HALL TEXTS IN STATISTICAL SCIENCE SERIES

Editors:

Dr Chris Chatfield
Reader in Statistics
School of Mathematical Sciences
University of Bath, UK

Professor Jim V. Zidek
Department of Statistics
University of British Columbia
Canada

OTHER TITLES IN THE SERIES INCLUDE

Practical Statistics for Medical Research
D.G. Altman

Interpreting Data
A.J.B. Anderson

Statistical Methods for SPC and TQM
D. Bissell

Statistics in Research and Development
Second edition
R. Caulcutt

The Analysis of Time Series
Fourth edition
C. Chatfield

Statistics in Engineering – A Practical Approach
A.V. Metcalfe

Statistics for Technology
Third edition
C. Chatfield

Introduction to Multivariate Analysis
C. Chatfield and A.J. Collins

Modelling Binary Data
D. Collett

Modelling Survival Data in Medical Research
D. Collett

Applied Statistics
D.R. Cox and E.J. Snell

Statistical Analysis of Reliability Data
M.J. Crowder, A.C. Kimber,
T.J. Sweeting and R.L. Smith

An Introduction to Generalized Linear Models
A.J. Dobson

Introduction to Optimization Methods and their Applications in Statistics
B.S. Everitt

Multivariate Studies – A Practical Approach
B. Flury and H. Riedwyl

Readings in Decision Analysis
S. French

Multivariate Analysis of Variance and Repeated Measures
D.J. Hand and C.C. Taylor

The Theory of Linear Models
B. Jorgensen

Statistical Theory
Fourth edition
B. Lindgren

Randomization and Monte Carlo Methods in Biology
B.F.J. Manly

Statistical Methods in Agriculture and Experimental Biology
Second edition
R. Mead, R.N. Curnow and A.M. Hasted

Elements of Simulation
B.J.T. Morgan

Probability: Methods and Measurement
A. O'Hagan

Essential Statistics
Third edition
D.G. Rees

Large Sample Methods in Statistics
P.K. Sen and J.M. Singer

Decision Analysis: A Bayesian Approach
J.Q. Smith

Applied Nonparametric Statistical Methods
Second edition
P. Sprent

Elementary Applications of Probability Theory
Second edition
H.C. Tuckwell

Statistical Process Control: Theory and Practice
Third edition
G.B. Wetherill and D.W. Brown

Statistics for Accountants
S. Letchford

Full information on the complete range of Chapman & Hall statistics books is available from the publishers.

BAYESIAN DATA ANALYSIS

Andrew B. Gelman
Columbia University,
New York, USA.

John S. Carlin
Royal Children's Hospital and
University of Melbourne,
Melbourne, Australia

Hal S. Stern
Iowa State University
Ames, USA

and

Donald B. Rubin
Harvard University
Cambridge, USA

CHAPMAN & HALL/CRC
Boca Raton London

Chapman & Hall/CRC Texts in Statistical Science Series

Bayesian Data Analysis

Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin

Bayesian Data Analysis is a comprehensive treatment of the statistical analysis of data from a Bayesian perspective. Modern computational tools are emphasized, and inferences are typically obtained using computer simulations.

The principles of Bayesian analysis are described with an emphasis on practical rather than theoretical issues, and illustrated using actual data. A variety of models are considered, including linear regression, hierarchical (random effects) models, robust models, generalized linear models and mixture models.

Two important and unique features of this text are thorough discussions of the methods for checking Bayesian models and the role of the design of data collection in influencing Bayesian statistical analysis.

Issues of data collection, model formulation, computation, model checking and sensitivity analysis are all considered. The student or practising statistician will find that there is guidance on all aspects of Bayesian data analysis.

Andrew Gelman, Columbia University, New York, USA. **John Carlin**, Royal Children's Hospital and University of Melbourne, Melbourne, Australia. **Hal Stern**, Iowa State University, Ames, USA. **Donald Rubin**, Harvard University, Cambridge, USA.

"*Bayesian Data Analysis* is easily the most comprehensive, scholarly, and thoughtful book on the subject, and I think will do much to promote the use of Bayesian methods"
— *Prof. David Blackwell, Dept. of Statistics, Univ. of California, Berkeley*

"[The book is] likely to have a profound influence on the direction of applied statistical work."
— *Prof. Roderick Little, Chairman, Dept. of Biostatistics, Univ. of Michigan*

"[The book is] a major contribution to the literature and is destined, in my opinion, to become a central text in the field".
— *Prof. Mike West, Director, Institute of Statistics and Decision Sciences, Duke University*

"a tour de force ... it is far more than an introductory text, and could act as a companion for a working scientist from undergraduate level through to professional life."
— *Robert Matthews, Aston University, in New Scientist, 28 September 1996*

C3991

ISBN 0-412-03991-5

90000

9 780412 039911

CHAPMAN & HALL/CRC

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress.

First edition 1995
Reprinted 1996, 1997, 1998
First CRC reprint 2000
Originally published by Chapman & Hall

Reprinted No claim to original U.S. Government works
International Standard Book Number 0-412-03991-5
Printed in the United States of America 3 4 5 6 7 8 9 0
Printed on acid-free paper

Contents

List of models

16.3. We are left with an improved model that still shows some lack of fit, suggesting possible directions for improved modeling and data collection.

16.5 Bibliographic note

Application of EM to mixture models is described in Dempster, Laird, and Rubin (1977). Gelman and King (1990b) fit a hierarchical mixture model using the Gibbs sampler in an analysis of elections, using an informative prior distribution to identify the mixture components separately. Other Bayesian applications of mixture models include Box and Tiao (1968), Turner and West (1993), and Belin and Rubin (1995b). Rubin and Stern (1994) and Gelman, Meng, and Stern (1996) demonstrate the use of posterior predictive checks to determine the number of mixture components required for an accurate model fit in an example in psychology. A comprehensive text emphasizing non-Bayesian approaches to finite mixture models is Titterington, Smith, and Makov (1985). West (1992) provides a brief review from a Bayesian perspective.

The schizophrenia example is discussed more completely in Belin and Rubin (1990, 1995a) and Gelman and Rubin (1992b). The posterior predictive checks for this example are presented in a slightly different graphical form in Gelman and Meng (1995). An expanded model applied to more complex data from schizophrenics appears in Rubin and Wu (1995).

CHAPTER 17

Models for missing data

17.1 Introduction

Our discussions of probability models in previous chapters, with few exceptions, assume that the desired dataset is completely observed. In this chapter we consider probability models and Bayesian methods for data analysis in problems with missing data. This chapter applies some of the terminology and notation of Chapter 7, which describes aspects of data collection that affect Bayesian data analysis, including mechanisms that lead to missing data.

We consider the two main tasks of data analysis with missing data as (1) multiple imputation—that is, simulating draws from the posterior predictive distribution of unobserved y_{mis} conditional on observed values y_{obs}, and (2) drawing from the posterior distribution of model parameters θ. The general idea is to extend the model specification to incorporate the missing observations and then to perform inference by averaging over the distribution of the missing values. In this chapter, we give a quick overview of theory and methods for multiple imputation and missing-data analysis; see the references at the end of the chapter for recent books on missing-data analysis from our Bayesian perspective.

17.2 Notation for data collection in the context of missing data problems

We begin by reviewing some notation from Chapter 7, focusing on the problem of unintentional missing data. As in Chapter 7, let y represent the 'complete data' that would be observed in the absence of missing values. The notation is intended to be quite general; y may be a vector of univariate measures or a matrix with each row containing the multivariate response variables of a single unit. Furthermore, it may be convenient to think of the complete data y as incorporating covariates, for example using a multivariate normal model for the vector of predictors and outcomes jointly in a regression context. We write $y = (y_{\text{obs}}, y_{\text{mis}})$, where y_{obs} denotes the

observed values and y_{mis} denotes the missing values. We also include in the model a random variable indicating whether each component of y is observed or missing. The *inclusion indicator* I is a data structure of the same size as y with each element of I equal to 1 if the corresponding component of y is observed and 0 if it is missing; we assume that I is completely observed. In a sample survey, item nonresponse corresponds to $I_{ij} = 0$ for unit i and item j, and unit nonresponse corresponds to $I_{ij} = 0$ for unit i and all items j.

The joint distribution of (y, I), given parameters (θ, ϕ), can be written as

$$p(y, I|\theta, \phi) = p(y|\theta)p(I|y, \phi).$$

The conditional distribution of I given the complete dataset y, indexed by the unknown parameter ϕ, describes the missing-data mechanism. The observed information is (y_{obs}, I); the distribution of the observed data is obtained by integrating over the distribution of y_{mis}:

$$p(y_{\text{obs}}, I|\theta, \phi) = \int p(y_{\text{obs}}, y_{\text{mis}}|\theta)p(I|y_{\text{obs}}, y_{\text{mis}}, \phi)dy_{\text{mis}}. \qquad (17.1)$$

Missing data are said to be *missing at random* (MAR) if the distribution of the missing-data mechanism does not depend on the missing values,

$$p(I|y_{\text{obs}}, y_{\text{mis}}, \phi) = p(I|y_{\text{obs}}, \phi),$$

so that the distribution of the missing-data mechanism is permitted to depend on other observed values (including fully observed covariates) and parameters ϕ. Formally, missing at random only requires the evaluation of $p(I|y, \phi)$ at the observed values of y_{obs}, not all possible values of y_{obs}. Under MAR, the joint distribution (17.1) of y_{obs}, I can be written as

$$\begin{aligned} p(y_{\text{obs}}, I|\theta, \phi) &= p(I|y_{\text{obs}}, \phi) \int p(y_{\text{obs}}, y_{\text{mis}}|\theta)dy_{\text{mis}} \qquad (17.2) \\ &= p(I|y_{\text{obs}}, \phi)p(y_{\text{obs}}|\theta). \end{aligned}$$

If, in addition, the parameters governing the distribution of the missing data mechanism, ϕ, and the parameters of the probability model, θ, are distinct, in the sense of being independent in the prior distribution, then Bayesian inferences for the model parameters θ can be obtained by considering only the observed-data likelihood, $p(y_{\text{obs}}|\theta)$. In this case, the missing-data mechanism is said to be *ignorable*.

In addition to the terminology of the previous paragraph, we speak of data that are *observed at random* (OAR) as well as MAR if the distribution of the missing-data mechanism is completely independent of y:

$$p(I|y_{\text{obs}}, y_{\text{mis}}, \phi) = p(I|\phi). \qquad (17.3)$$

In such cases, we say the missing data are *missing completely at random* (MCAR). The preceding paragraph shows that the weaker pair of assumptions of MAR and distinct parameters is sufficient for obtaining Bayesian inferences without requiring further modeling of the missing-data mechanism. Since it is relatively rare in practical problems for MCAR to be plausible, we focus in this chapter on methods suitable for the more general case of MAR.

The plausibility of MAR (but not MCAR) is enhanced by including as many observed characteristics of each individual or object as possible when defining the dataset y; that is, y includes both x and y, in the notation of Chapter 7. Increasing the pool of observed variables (with relevant variables) decreases the degree to which missingness depends on unobservables given the observed variables.

We conclude this section with a discussion of several examples that illustrate the terminology and principles described above. Suppose that measurements consist of two variables $y = (\text{age, income})$ with age recorded for all individuals but income missing for some individuals. For simplicity of discussion, we model the joint distribution of the outcomes as bivariate normal. If the probability that income is recorded is the same for all individuals, independent of age and income, then the data are MAR and OAR, and therefore MCAR. If the probability that income is missing depends on the age of the respondent but not on the income of the respondent given age, then the data are MAR but not OAR. The missing-data mechanism is ignorable when, in addition to MAR, the parameters governing the missing-data process are distinct from those of the bivariate normal distribution (as is typically the case with standard models). If, as seems likely, the probability that income is missing depends on age group and moreover on the value of income within each age group, then the data are neither MAR nor OAR. The missing data mechanism in this last case is said to be nonignorable.

The relevance of the missing-data mechanism depends on the goals of the data analysis. If we are only interested in the mean and variance of the age variable, then we can discard all recorded income data and construct a model in which the missing-data mechanism is ignorable. On the other hand, if we are interested in the marginal distribution of income, then the missing-data mechanism is of paramount importance and must be carefully considered.

If information about the missing-data mechanism is available, then it may be possible to perform an appropriate analysis even if the missing-data mechanism is nonignorable, as discussed in Chapter 7.

17.3 Computation and multiple imputation

Bayesian computation in a missing-data problem is based on the joint posterior distribution of parameters and missing data, given modeling assumptions and observed data. The result of the computation is a set of vectors

of simulations of all unknowns, $(y^l_{\text{mis}}, \theta^l), l = 1, \ldots, L$. At this point, there are two possible courses of action:

- Obtain inferences for any parameters, missing data, and predictive quantities of interest.

- Report the results in the form of the observed data and the simulated vectors y^l_{mis}, which are called *multiple imputations*. Other users of the data can then use these multiply imputed complete datasets and perform analysis without needing to model the missing-data mechanism.

In the context of this book, the first option seems most natural, but in practice, especially when most of the data values are *not* missing, it is generally useful to divide a data analysis in two parts: first, cleaning the data and multiply imputing missing values, and second, performing inference about quantities of interest using the imputed datasets.

In either case, computation usually starts by crude methods of imputation based on approximate models such as MCAR. The initial imputations are used as starting points for iterative mode-finding and simulation algorithms as discussed in Part III of this book.

Application of the EM algorithm to find modes of the observed-data posterior density

The EM algorithm was described in some detail in Chapter 9 as an approach for obtaining the posterior mode in complex problems. As was mentioned there, the EM algorithm formalizes a fairly old approach to handling missing data: replace missing data by estimated values, estimate model parameters, and perhaps, repeat these two steps several times. Often, a problem with no missing data can be easier to analyze if the dataset is augmented by some unobserved values, which may be thought of as missing data. Thus the EM algorithm is extremely useful, even in many problems that would not usually be considered missing-data problems.

Here, we briefly review the EM algorithm and its extensions using the notation of this chapter. The EM algorithm can be applied whether the missing data are ignorable or not by including the missing-data model in the likelihood. For ease of exposition, we assume the missing-data mechanism is ignorable and therefore omit the inclusion indicator I in the following explanation. The generalization to specified nonignorable models is relatively straightforward. We assume that any augmented data, for example, mixture component indicators, are included as part of y_{mis}. Converting to the notation of Section 9.5:

Notation in Section 9.5	Notation for missing data
Data, y	Observed information (y_{obs}, I)
Marginal mode of parameters ϕ	Posterior mode of parameters (θ or, if estimating the missingness mechanism, (θ, ϕ))
Averaging over parameters γ	Averaging over missing data, y_{mis}

The EM algorithm is best known as it is applied to exponential families. In that case, the expected complete-data log posterior density is linear in the expected complete-data sufficient statistics so that only the latter need be evaluated or imputed. Examples are provided in the next section.

EM for nonignorable models. For a nonignorable missingness mechanism the EM algorithm can also be applied as long as a model for the missing data is specified (for example, censored or rounded data with known censoring point or rounding rule). The only change in the EM algorithm is that all calculations explicitly condition on the inclusion indicator I. Specifically, the expected complete-data log posterior density is a function of model parameters θ and missing-data-mechanism parameters ϕ, conditional on the observed data y_{obs} and the inclusion indicator I, averaged over the distribution of y_{mis} at the current values of the parameters $(\theta^{\text{old}}, \phi^{\text{old}})$.

Computational short-cut with monotone missing-data patterns. A dataset is said to have a *monotone pattern of missing data* if the variables can be ordered in blocks such that the first block of variables is more observed than the second block of variables (that is, values in the first block are present whenever values in the second are present but the converse does not necessarily hold), the second block of variables is more observed than the third block, and so forth. Many datasets have this pattern or nearly so. Obtaining posterior modes can be especially easy when the data have a monotone pattern. For instance, with normal data, rather than computing a separate regression estimate conditioning on y_{obs} in the E-step for each observation, the monotone pattern implies that there are only as many patterns of missing data as there are blocks of variables. Thus, all of the observations with the same pattern of missing data can be handled in a single step. For data that are close to the monotone pattern, the EM algorithm can be applied as a combination of two approaches: first, the E-step can be carried out for those values of y_{mis} that are outside the monotone pattern; then, the more efficient calculations can be carried out for the missing data that are consistent with the monotone pattern.

17.4 Missing data in the multivariate normal and t models

We consider the basic continuous-data model in which y represents a sample of size n from a d-dimensional multivariate normal distribution $\text{N}_d(\mu, \Sigma)$

with y_{obs} the set of observed values and y_{mis} the set of missing values. We assume a uniform prior distribution for simplicity; it is straightforward to generalize to a conjugate or hierarchical prior distribution.

Finding posterior modes using EM

The multivariate normal is an exponential family with sufficient statistics equal to

$$\sum_{i=1}^{n} y_{ij}, \quad j = 1, \ldots, d$$

and

$$\sum_{i=1}^{n} y_{ij} y_{ik}, \quad j, k = 1, \ldots, d.$$

Let $y_{obs\,i}$ denote the components of $y_i = (y_{i1}, \ldots, y_{id})$ that are observed and $y_{mis\,i}$ denote the missing components. Let $\theta^{old} = (\mu^{old}, \Sigma^{old})$ denote the current estimates of the model parameters. The E step of the EM algorithm computes the expected value of these sufficient statistics conditional on the observed values and the current parameter estimates. Specifically,

$$\begin{aligned} \mathrm{E}\left(\sum_{i=1}^{n} y_{ij} | y_{obs}, \theta^{old}\right) &= \sum_{i=1}^{n} y_{ij}^{old} \\ \mathrm{E}\left(\sum_{i=1}^{n} y_{ij} y_{ik} | y_{obs}, \theta^{old}\right) &= \sum_{i=1}^{n} \left(y_{ij}^{old} y_{ik}^{old} + c_{ijk}^{old}\right), \end{aligned}$$

where

$$y_{ij}^{old} = \begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed} \\ \mathrm{E}(y_{ij} | y_{obs}, \theta^{old}) & \text{if } y_{ij} \text{ is missing,} \end{cases}$$

and

$$c_{ijk}^{old} = \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ are observed} \\ \mathrm{cov}(y_{ij}, y_{ik} | y_{obs}, \theta^{old}), & \text{if } y_{ij} \text{ and } y_{ik} \text{ are missing.} \end{cases}$$

The conditional expectation and covariance are easy to compute: the conditional posterior distribution of the missing elements of y_i, $y_{mis\,i}$, given y_{obs} and θ, is multivariate normal with mean vector and variance matrix obtained from the full mean vector and variance matrix θ^{old} as in Appendix A.

The M-step of the EM algorithm uses the expected complete-data sufficient statistics to compute the next iterate, θ^{new}. Specifically,

$$\mu_j^{new} = \frac{1}{n} \sum_{i=1}^{n} y_{ij}^{old}, \text{ for } j = 1, \ldots, d,$$

and

$$\sigma_{jk}^{new} = \frac{1}{n} \sum_{i=1}^{n} (y_{ij}^{old} y_{ik}^{old} + c_{ijk}^{old}) - \mu_j^{new} \mu_k^{new}, \text{ for } j, k = 1, \ldots, d.$$

Starting values for the EM algorithm can be obtained using crude methods As always when finding posterior modes, it is wise to use several starting values in case of multiple modes. It is crucial that the initial estimate of the variance matrix be positive definite; thus various estimates based on complete cases (that is, units with all outcomes observed), if available, can be useful.

Drawing samples from the posterior distribution of the model paramete

One can draw imputations for the missing values from the normal model using the modal estimates as starting points for data augmentation (the Gibbs sampler) on the joint posterior distribution of missing values and parameters, alternately drawing y_{mis}, μ, and Σ from their conditional posterior distributions. For more complicated models, some of the steps of the Gibbs sampler must be replaced by Metropolis steps.

As with the EM algorithm, considerable gains in efficiency are possible if the missing data have a monotone pattern. In fact, for monotone missing data, it is possible under an appropriate parameterization to draw directly from the incomplete-data posterior distribution, $p(\theta | y_{obs})$. Suppose that y_1 is more observed than y_2, y_2 is more observed than y_3, and so forth. To be specific, let $\psi = \psi(\theta) = (\psi_1, \ldots, \psi_k)$, where ψ_1 denotes the parameters of the marginal distribution of the first block of variables in the monotone pattern y_1, ψ_2 denotes the parameters of the conditional distribution of y_2 given y_1, and so on (the normal distribution is d-dimensional but the monotone pattern is defined by k blocks of variables). For multivari normal data, ψ_j contains the parameters of the linear regression of y_j on $y_1, \ldots, y_{j-1}$—the regression coefficients and the residual variance matrix. The parameter ψ is a one-to-one function of the parameter θ, and the complete parameter space of ψ is the product of the parameter spaces of $\psi_1, \ldots, \psi_k$. The likelihood factors into k distinct pieces, so that

$$\log p(\psi | y_{obs}) = \log p(\psi_1 | y_{obs}) + \log p(\psi_2 | y_{obs}) + \ldots + \log p(\psi_k | y_{obs}),$$

with the jth piece depending only on the parameters ψ_j. If a prior distribution $p(\psi)$ factors as

$$p(\psi) = p(\psi_1) p(\psi_2 | \psi_1) \cdots p(\psi_k | \psi_1, \ldots, \psi_{k-1}),$$

then it is possible to draw directly from the posterior distribution in sequence: first draw ψ_1, then ψ_2 conditional on ψ_1 and y_{obs}, and so forth.

For a missing-data pattern that is not precisely monotone, we can define a monotone data augmentation algorithm that imputes only enough

data to obtain a monotone pattern. The imputation step draws a sample from the conditional distribution of the elements in y_{mis} that are needed to create a monotone pattern. The posterior step then draws directly from the posterior distribution taking advantage of the monotone pattern. Typically, the monotone data augmentation algorithm will be more efficient than ordinary data augmentation if the departure from monotonicity is not substantial, because fewer imputations are being done and analytic calculations are being used to replace the other simulation steps. There may be several ways to order the variables that each lead to nearly monotone patterns. Determining the best such choice is complicated since 'best' is defined by providing the fastest convergence of an iterative simulation method. One simple approach is to choose y_1 to be the variable with the fewest missing values, y_2 to be the variable with the second fewest, and so on.

Student-t extensions of the normal model

Chapter 12 describes robust alternatives to the normal model based on the Student-t distribution. Such models can be useful for accommodating data prone to outliers, or as a means of performing a sensitivity analysis on a normal model. Suppose now that the intended data consist of multivariate observations,

$$y_i|\theta, V_i \sim \mathrm{N}_d(\mu, V_i\Sigma),$$

where V_i are unobserved iid random variables with an Inv-$\chi^2(\nu, 1)$ distribution. For simplicity, we consider ν to be specified; if unknown, it is another parameter to be estimated.

Data augmentation can be applied to the t model with missing values in y_i by adding a step that imputes values for the V_i, which are thought of as additional missing data. The imputation step of data augmentation consists of two parts. First, a value is imputed for each V_i from its posterior distribution given y_{obs}, θ, ν. This posterior distribution is a product of the normal distribution for $y_{obs\,i}$ given V_i and the scaled inverse-χ^2 prior distribution for V_i,

$$p(V_i|y_{obs}, \theta, \nu) \propto \mathrm{N}(y_{obs\,i}|\mu_{obs\,i}, \Sigma_{obs\,i}V_i)\text{Inv-}\chi^2(V_i|\nu, 1), \qquad (17.4)$$

where $\mu_{obs\,i}, \Sigma_{obs\,i}$ refer to the elements of the mean vector and variance matrix corresponding to components of y_i that are observed. The conditional posterior distribution (17.4) is easily recognized as scaled inverse-χ^2, so obtaining imputed values for V_i is straightforward. The second part of each iteration step is to impute the missing values $y_{mis\,i}$ given (y_{obs}, θ, V_i), which is identical to the imputation step for the ordinary normal model since given V_i, the value of $y_{mis\,i}$ is obtained as a draw from the conditional normal distribution. The posterior step of the data augmentation algorithm

treats the imputed values as if they were observed and is, therefore, a complete-data weighted multivariate normal problem. The complexity of th step depends on the prior distribution for θ.

The E-step of the EM algorithm for the t extensions of the normal mod is obtained by replacing the imputation steps above with expectation step Thus the conditional expectation of V_i from its scaled inverse-χ^2 posterio distribution and conditional means and variances of $y_{mis\,i}$ would be used i place of random draws. The M-step finds the conditional posterior mod rather than sampling from the posterior distribution. When the degrees o freedom parameter for the t distribution is allowed to vary, the ECM and ECME algorithms can be used.

Nonignorable models. The principles for performing Bayesian inferenc nonignorable models based on the normal or t distributions are analogous to those presented in the ignorable case. At each stage, θ is supplemented by any parameters of the missing-data mechanism, ϕ, and inference is conditional on observed data y_{obs} and the inclusion indicator I.

17.5 Missing values with counted data

The analysis of fully observed counted data is discussed in Section 3.5 for saturated multinomial models and in Chapter 14 for loglinear models. Here, we consider how those techniques can be applied to missing discrete data problems.

Multinomial samples. Suppose that the hypothetical complete data are a multinomial sample of size n with cells $c_1, \ldots, c_J$, cell probabilities $\theta = (\pi_1, \ldots, \pi_J)$, and cell counts $n_1, \ldots, n_J$. Conjugate prior distributions for θ are in the Dirichlet family (see Appendix A and Section 3.5). The observed data are m completely classified observations with m_j in the jth cell, and $n - m$ partially classified observations (the missing data). The partially classified observations are known to fall in subsets of the J cells. For example, in a $2 \times 2 \times 2$ table, $J = 8$, and an observation with known classification for the first two dimensions but with missing classification for the third dimension is known to fall in one of two possible cells.

Partially classified observations. It is convenient to organize each of the $n - m$ partially classified observations according to the subset of cells to which it can belong. Thus suppose there are K types of partially classified observation, and the r_k observations of the kth type are known to fall in one of the cells in subset S_k.

The iterative procedure used for normal data in the previous subsection, data augmentation, can be used here to iterate between imputing cells for the partially classified observations and obtaining draws from the posterior distribution of the parameters θ. The imputation step draws from the conditional distribution of the partially classified cells given the observed data

and the current set of parameters θ. For each $k = 1, \ldots, K$, the r_k partially classified observations known to fall in the subset of cells S_k are assigned randomly to each of the cells in S_k with probability

$$\frac{\pi_j I_{j \in S_k}}{\sum_{l=1}^{J} \pi_l I_{l \in S_k}},$$

where $I_{j \in S_k}$ is the indicator function equal to 1 if cell j is part of the subset S_k and 0 otherwise. When the prior distribution is Dirichlet, then the posterior step requires drawing from the conjugate Dirichlet posterior distribution, treating the imputed data and the observed data as a complete dataset. As usual, it is possible to use other, nonconjugate, prior distributions, although this makes the posterior computation more difficult. The EM algorithm for exploring modes is a nonstochastic version: the E-step computes the number of the r_k partially classified observations that are expected to fall in each cell (the mean of the multinomial distribution), and the M-step computes updated cell probability estimates by combining the observed cell counts with the results of the E-step.

As usual, the analysis is simplified when the missing-data pattern is monotone or nearly monotone, so that the likelihood can be written as a product of the marginal distribution of the most observed set of variables and a set of conditional distributions for each subsequent set of variables conditional on all of the preceding, more observed variables. If the prior density is also factored, for example as a product of Dirichlet densities for the parameters of each factor in the likelihood, then the posterior distribution can be drawn from directly. The analysis of nearly monotone data requires iterating two steps: imputing values for those partially classified observations required to complete the monotone pattern, and drawing from the posterior distribution, which can be done directly for the monotone pattern.

Further complications arise when the cell probabilities θ are modeled, as in loglinear models; see the references at the end of this chapter.

17.6 Example: an opinion poll in Slovenia

We illustrate the methods described in the previous section with the analysis of an opinion poll concerning independence in Slovenia, formerly a province of Yugoslavia and now a nation. In 1990, a plebiscite was held in Slovenia at which the adult citizens voted on the question of independence. The rules of the plebiscite were such that nonattendance, as determined by an independent and accurate census, was equivalent to voting 'no'; only those attending and voting 'yes' would be counted as being in favor of independence. In anticipation of the plebiscite, a Slovenian public opinion survey had been conducted that included several questions concerning likely plebiscite attendance and voting. In that survey, 2074 Slovenians were

Secession	Attendance	Independence: Yes	Independence: No	Independence: Don't know
Yes	Yes	1191	8	21
	No	8	0	4
	Don't Know	107	3	9
No	Yes	158	68	29
	No	7	14	3
	Don't Know	18	43	31
Don't Know	Yes	90	2	109
	No	1	2	25
	Don't Know	19	8	96

Table 17.1. $3 \times 3 \times 3$ *table of results of 1990 pre-plebiscite survey in Slovenia, from Rubin, Stern, and Vehovar (1995). We treat 'don't know' responses as missing data. Of most interest is the proportion of the electorate whose 'true' answers are 'yes' on both 'independence' and 'attendance.'*

asked three questions: (1) Are you in favor of independence?, (2) Are you in favor of secession?, and (3) Will you attend the plebiscite? The results of the survey are displayed in Table 17.1. Let α represent the estimand of interest from the sample survey, the proportion of the population planning to attend and vote in favor of independence. It follows from the rules of the plebiscite that 'don't know' (DK) can be viewed as missing data (at least accepting that 'yes' and 'no' responses to the survey are accurate for the plebiscite). Every survey participant will vote yes or no—perhaps directly or perhaps indirectly by not attending.

Why ask three questions when we only care about two of them? The response to question 2 is not directly relevant but helps us more accurately impute the missing data. The survey participants may provide some information about their intentions by their answers to question 2; for example, a 'yes' response to question 2 might be indicative of likely support for independence for a person who did not answer question 1.

Crude estimates

As an initial estimate of α, the proportion planning to attend and vote 'yes,' we ignore the DK responses for these two questions; considering only the 'available cases' (those answering the attendance and independence questions) yields a crude estimate $\hat{\alpha} = 1439/1549 = 0.929$, which seems to suggest that the outcome of the plebiscite is not in doubt. However, given that only 1439/2074, or 69%, of the survey participants definitely plan to attend and vote 'yes,' and given the importance of the outcome,

improved inference is desirable, especially considering that if we were to assume that the DK responses correspond to 'no,' we would obtain a very different estimate.

The 2074 responses include those of substitutes for original survey participants who could not be contacted after several attempts. Although the substitutes were chosen from the same clusters as the original participants to minimize differences between substitutes and nonsubstitutes, there may be some concern about differences between the two groups. We indicate the most pessimistic estimate for α by noting that only 1251/2074 of the original survey sample (using information not included in the table) plans to attend and vote 'yes.' For simplicity, we treat substitutes as original respondents for the remainder of this section and ignore the effects of clustering.

The likelihood and prior distribution

The complete data can be viewed as a sample of 2074 observations from a multinomial distribution on the eight cells of the $2 \times 2 \times 2$ contingency table, with corresponding vector of probabilities θ; the DK responses are treated as missing data. We use θ_{ijk} to indicate the probability of the multinomial cell in which the respondent gave answer i to question 1, answer j to question 2, and answer k to question 3, with $i, j, k = 0$ for 'no' and 1 for 'yes.' The estimand of most interest, α, is the sum of the appropriate elements of θ, $\alpha = \theta_{101} + \theta_{111}$. In our general notation, y_{obs} are the observed 'yes' and 'no' responses, and y_{mis} are the 'true' 'yes' and 'no' responses corresponding to the DK responses. The 'complete data' form a 2074×3 matrix of 0's and 1's that can be recoded as a contingency table of 2074 counts in eight cells.

The Dirichlet prior distribution for θ with parameters all equal to zero is noninformative in the sense that the posterior mode is the maximum likelihood estimate. Since one of the observed cell counts is 0 ('yes' on secession, 'no' on attendance, 'no' on independence), the improper prior distribution does not lead to a proper posterior distribution. It would be possible to proceed with the improper prior density if we thought of this cell as being a structural zero—a cell for which a nonzero count is impossible. The assumption of a structural zero does not seem particularly plausible here, and we choose to use a Dirichlet distribution with parameters all equal to 0.1 as a convenient (though arbitrary) way of obtaining a proper posterior distribution while retaining a diffuse prior distribution. A thorough analysis should explore the sensitivity of conclusions to this choice of prior distribution.

The model for the 'missing data'

We treat the DK responses as *missing values*, each known only to belong to some subset of the eight cells. Let $n = (n_{ijk})$ represent the hypothetical complete data and let $m = (m_{ijk})$ represent the number of completely classified respondents in each cell. There are 18 types of partially classified observations; for example, those answering 'yes' to questions 1 and 2 and DK to question 3, those answering 'no' to question 1 and DK to questions 2 and 3, and so on. Let r_p denote the number of partially classified observations of type p; for example, let r_1 represent the number of those answering 'yes' to questions 1 and 2 and DK to question 3. Let S_p denote the set of cells to which the partially classified observations of the pth type might belong; for example, S_1 includes the 111 and 110 cells. We assume that the DK responses are missing at random, which implies that the probability of a DK response may depend on the answers to other questions but not on the unobserved response to the question at hand.

The complete-data likelihood is

$$p(n|\theta) \propto \prod_{i=0}^{1} \prod_{j=0}^{1} \prod_{k=0}^{1} \theta_{ijk}^{n_{ijk}}$$

with complete-data sufficient statistics $n = (n_{ijk})$. If we let $\pi_{ijk\,p}$ represent the probability that a partially classified observation of the pth type belongs in cell ijk, then the MAR model implies that, given a set of parameter values θ, the distribution of the r_p observations with the pth missing-data pattern is multinomial with probabilities

$$\pi_{ijk\,p} = \frac{\theta_{ijk} I_{ijk \in S_p}}{\sum_{i'j'k'} \theta_{i'j'k'} I_{i'j'k' \in S_p}},$$

where the indicator $I_{ijk \in S_p}$ is 1 if cell ijk is in subset S_p and 0 otherwise.

Using the EM algorithm to find the posterior mode of θ

The EM algorithm in this case finds the mode of the posterior distribution of the multinomial probability vector θ by averaging over the missing data (the DK responses) and is especially easy here with the assumption of distinct parameters. For the multinomial distribution, the E-step, computing the expected complete-data log posterior density, is equivalent to computing the expected counts in each cell of the contingency table given the current parameter estimates. The expected count in each cell consists of the fully observed cases in the cell and the expected number of the partially observed cases that fall in the cell. Under the missing at random assumption and the resulting multinomial distribution, the DKs in each category (that is, the pattern of missing data) are allocated to the possible cells in proportion to the current estimate of the model parameters. Mathematically,

given current parameter estimate θ^{old}, the E-step computes

$$n_{ijk}^{\text{old}} = \text{E}(n_{ijk}|m, r, \theta^{\text{old}}) = m_{ijk} + \sum_p r_p \pi_{ijkp}.$$

The M-step computes new parameter estimates based on the latest estimates of the expected counts in each cell; for the saturated multinomial model here (with distinct parameters), $\theta^{\text{new}} = (n_{ijk}^{\text{old}} + 0.1)/(n + 0.8)$. The EM algorithm is considered to converge when none of the parameter estimates changes by more than a tolerance criterion, which we set here to the unnecessarily low value of 10^{-16}. The posterior mode of α is 0.882, which turned out to be quite close to the eventual outcome in the plebiscite.

Using SEM to estimate the posterior variance matrix and obtain a normal approximation

To complete the normal approximation, estimates of the posterior variance matrix are required. The SEM algorithm numerically computes estimates of the variance matrix using the EM program and the complete-data variance matrix (which is available since the complete data are modeled as multinomial).

The SEM algorithm is applied to the logit transformation of the components of θ, since the normal approximation is generally more accurate on this scale. Posterior central 95% intervals for logit(α) are transformed back to yield a 95% interval for α, $[0.857, 0.903]$. The standard error was inflated to account for the design effect of the clustered sampling design using approximate methods based on the normal distribution.

Multiple imputation using data augmentation

Even though the sample size is large in this problem, it seems prudent, given the missing data, to perform posterior inference that does not rely on the asymptotic normality of the maximum likelihood estimates. The data augmentation algorithm, a special case of the Gibbs sampler, can be used to obtain draws from the posterior distribution of the cell probabilities θ under a noninformative Dirichlet prior distribution. As described earlier, for count data, the data augmentation algorithm iterates between imputations and posterior draws. At each imputation step, the r_p cases with the pth missing-data pattern are allocated among the possible cells as a draw from a multinomial distribution. Conditional on these imputations, a draw from the posterior distribution of θ is obtained from the Dirichlet posterior distribution. A total of 1000 draws from the posterior distribution of θ were obtained—the second half of 20 data augmentation series, each run for 100 iterations, at which point the potential scale reductions, $\sqrt{\tilde{R}}$, were below 1.1 for all parameters.

Posterior inference for the estimand of interest

The posterior median of α, the population proportion planning to attend and vote yes, is 0.883. We construct an approximate posterior central 95% interval for α by inflating the variance of the 95% interval from the posterior simulations to account for the clustering in the design (to avoid the complications but approximate the results of a full Bayesian analysis of this sampling design); the resulting interval is $[0.859, 0.904]$. It is not surprising, given the large sample size, that this interval matches the interval obtained from the asymptotic normal distribution.

By comparison, neither of the initial crude calculations is very close to the actual plebiscite outcome, in which 88.5% of the eligible population attended and voted 'yes.'

17.7 Inference using multiple imputation

Imputation

For complex surveys, the modeling approach to obtaining posterior inference when some data values are missing can be carried out separately by each user. For surveys with many possible users, it is desirable to provide an easier method. A standard practice in complex surveys is to replace each of the missing values by an imputed value. A single imputation provides a complete dataset that can be used by a variety of researchers to address a variety of questions. Assuming the imputation model is reasonable, the results from an analysis of the imputed dataset are likely to provide more accurate estimates than would be obtained by discarding data with missing values. In addition, it is likely that the creator of the imputation will have access to many more variables than the ultimate users and can therefore create superior imputations. Two limitations are that (1) the quality of the ultimate analysis depends on the quality of the imputation, and (2) the single imputation does not address the sampling variability under the nonresponse model, thus leading to falsely precise inferences.

Multiple imputation

The key idea of multiple imputation is to create more than one set of imputations. This addresses one of the difficulties of single imputation in that the uncertainty due to nonresponse under a particular missing-data model can be properly reflected. The data augmentation algorithm that is used in this chapter to obtain posterior inference can be viewed as iterative multiple imputation. Here we restrict attention to the context of complex surveys for which a relatively small number of multiple imputations can be used to investigate the variability of the missing-data model. We give some simple approximations that are widely applicable. To be specific, if there

are K sets of imputed values under a single model, let $\hat{\theta}_k, \widehat{W}_k, k = 1, \ldots, K$, be the K complete-data parameter estimates and associated variance estimates for the scalar parameter θ. The K complete-data analyses can be combined to form the combined estimate of θ,

$$\bar{\theta}_K = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}_k.$$

The variability associated with this estimate has two components: the average of the complete-data variances (the within-imputation component),

$$\overline{W}_K = \frac{1}{K} \sum_{k=1}^{K} \widehat{W}_k,$$

and the variance across the different imputations (the between-imputation component),

$$B_K = \frac{1}{K-1} \sum_{k=1}^{K} (\hat{\theta}_k - \bar{\theta}_K)^2.$$

The total variability associated with $\bar{\theta}_K$ is

$$T_K = \overline{W}_K + \frac{K+1}{K} B_K.$$

The reference distribution for creating interval estimates for θ is a Student-t distribution with approximate degrees of freedom based on a Satterthwaite approximation,

$$\text{d.f.} = (K-1)\left(1 + \frac{1}{K+1} \frac{\overline{W}_K}{B_K}\right)^2.$$

If the fraction of missing information is not too high, then posterior inference will likely not be sensitive to modeling assumptions about the missing-data mechanism. One approach is to create a 'reasonable' missing-data model, and then check the sensitivity of the posterior inferences to other missing-data models. In particular, it often seems helpful to begin with an ignorable model and explore the sensitivity of posterior inferences to plausible nonignorable models.

17.8 Bibliographic note

The jargon 'missing at random,' 'observed at random,' and 'ignorability' originated with Rubin (1976). Factoring general distributions with monotone or nearly monotone missing data and the definition of 'distinct parameters' originated with Rubin (1974a, 1976) and is extended in Rubin (1987a); an early important example appears in Anderson (1957). Multiple imputation was proposed in Rubin (1978b) and is discussed in detail in

Rubin (1987a) with a focus on sample surveys; Rubin (1996) is a recent review of the topic. Kish (1965) and Madow et al. (1983) discuss less formal ways of handling missing data in sample surveys.

Meng (1994) discusses the theory of multiple imputation when different models are used for imputation and analysis. Clogg et al. (1991) and Belin et al. (1993) describe hierarchical logistic regression models used for imputation for the U.S. Census. There has been growing use of multiple imputation using nonignorable models for missing data; for example, Heitjan and Landis (1994) set up a model for unobserved medical outcomes and multiply impute using matching to appropriate observed cases. David et al. (1986) present a thorough discussion and comparison of a variety of imputation methods for a missing data problem in survey imputation.

Little and Rubin (1987) is a comprehensive text on statistical analysis with missing data. Tanner and Wong (1987) describe the use of data augmentation to calculate posterior distributions. Schafer (1997) applies data augmentation for multivariate exchangeable models, including the normal and loglinear models discussed briefly in this chapter; Liu (1995) extends these methods to t models. The Slovenia survey is described in more detail in Rubin, Stern, and Vehovar (1995).

17.9 Exercises

1. Computation for discrete missing data: reproduce the results of Section 17.6 for the 2×2 table involving independence and attendance. You can ignore the clustering in the survey and pretend it was obtained from a simple random sample. Specifically:
   (a) Use EM to obtain the posterior mode of α, the proportion who will attend and will vote 'yes.'
   (b) Use SEM to obtain the asymptotic posterior standard deviation of logit(α), and thereby obtain an approximate 95% interval for α.
   (c) Use Markov chain simulation of the parameters and missing data to obtain the approximate posterior distribution of θ. Clearly say what your starting distribution is for your simulations. Be sure to simulate more than one sequence and to include some diagnostics on the convergence of the sequences.
2. Monotone missing data: create a monotone pattern of missing data for the opinion poll data of Section 17.6 by discarding some observations. Compare the results of analyzing these data with the results given in that section.

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.